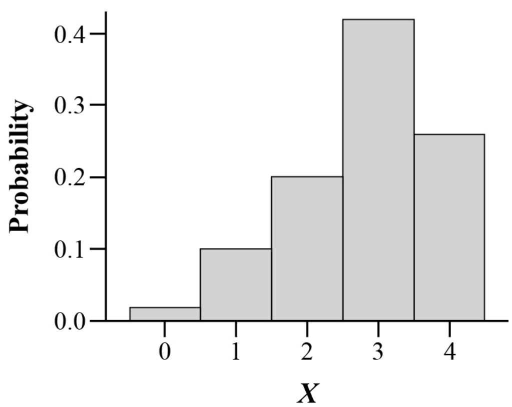
Chapter 6

Section 6.1

Check Your Understanding, page 350:

1. $P(X \geq 3)$ is the probability that the student got either an A or a B. This probability is $P(X \geq 3) = 0.42 + 0.26 = 0.68$.
2. We are looking for $P(X < 2) = 0.02 + 0.10 = 0.12$. There is a 0.12 probability that the student got worse than a C.
3. The histogram (shown below) is left skewed. This means that higher grades are more likely, but that there are a few lower grades. The median grade is a B ($X = 3$) and the grades vary from A ($X = 4$) to F ($X = 0$).



Check Your Understanding, page 355:

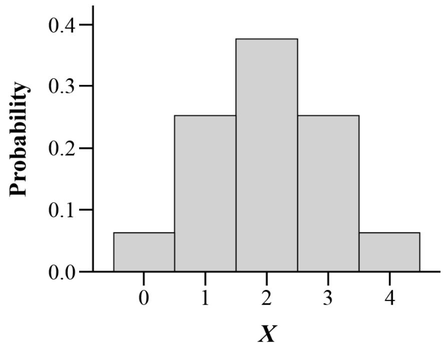
1. $\mu_X = 0(0.3) + 1(0.4) + 2(0.2) + 3(0.1) = 1.1$. If many, many Fridays are randomly selected, the average number of cars sold will be about 1.1.
2. $\sigma_X^2 = (0-1.1)^2(0.3) + (1-1.1)^2(0.4) + (2-1.1)^2(0.2) + (3-1.1)^2(0.1) = 0.89$. So $\sigma_X = \sqrt{0.89} = 0.943$. The number of cars sold on a randomly selected Friday will typically vary from the mean (1.1) by about 0.943 cars.

Exercises, page 359:

6.1 (a) If you toss a coin 4 times, the sample space is HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, HTTH, THHT, THTH, TTHH, HTTT, THTT, TTHT, TTTH, TTTT. All of these outcomes are equally likely and have probability $\frac{1}{16}$. To find the probability of X taking a specific value, count the number of outcomes with exactly this number of heads. Here is the probability distribution:

Value	0	1	2	3	4
Probability	1/16	4/16	6/16	4/16	1/16

(b) The histogram below shows that this distribution is symmetric with a center at 2.

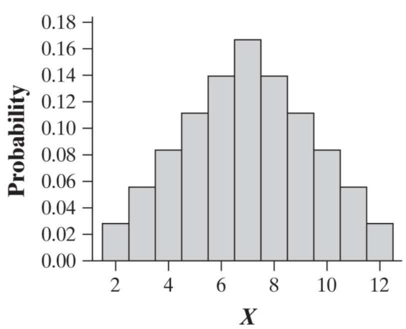


(c) $P(X \le 3) = 1 - P(X = 4) = 1 - 1/16 = 15/16 = 0.9375$. There is a 0.9375 probability that you will get three or fewer heads in 4 tosses of a fair coin.

6.2 (a) If we roll two 6-sided dice, the sample space is (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6). All of these are equally likely, so the probability of any one outcome is $\frac{1}{36}$. To find the probability of X taking a specific value, count the number of outcomes where the sum of the dice is exactly this number. Here is the probability distribution:

Value	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

(b) The histogram below shows that the distribution is symmetric with a center at 7.



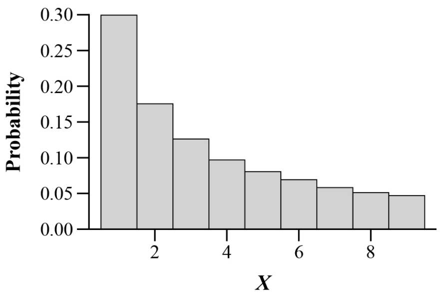
(c) $P(T \ge 5) = 1 - P(T \le 4) = 1 - \left(\frac{1}{36} + \frac{2}{36} + \frac{3}{36}\right) = 1 - \frac{6}{36} = \frac{30}{36} = 0.833$. There is a 0.833 probability that you will get a sum of 5 or more when you roll a pair of 6-sided dice.

6.3 (a) "At least one nonword error" is the event $X \ge 1$. $P(X \ge 1) = 1 - P(X = 0) = 1 - 0.1 = 0.9$.
(b) The event $X \le 2$ is "at most two nonword errors." $P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.1 + 0.2 + 0.3 = 0.6$. $P(X < 2) = P(X = 0) + P(X = 1) = 0.1 + 0.2 = 0.3$.

6.4 (a) "Plays with at most two toys" is the event $X \le 2$. $P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.03 + 0.16 + 0.30 = 0.49$.

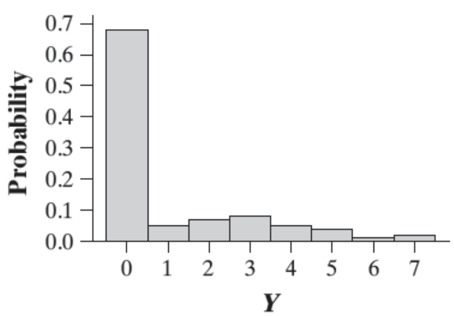
(b) The event $X > 3$ is "the child plays with more than three toys."
$P(X > 3) = P(X = 4) + P(X = 5) = 0.17 + 0.11 = 0.28.$ $P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.49 = 0.51.$

6.5 (a) All of the probabilities are between 0 and 1 and they sum to 1 so this is a legitimate probability distribution.
(b) The histogram below shows a right skewed distribution. The most likely first digit is 1 and each subsequent digit is less likely than the previous digit.



(c) The event $X \geq 6$ is the event that "the first digit in a randomly chosen record is a 6 or higher."
$P(X \geq 6) = P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) = 0.067 + 0.058 + 0.051 + 0.046 = 0.222.$
(d) The event that "the first digit is at most 5" is the event $X \leq 5$.
$P(X \leq 5) = 1 - P(X \geq 6) = 1 - 0.222 = 0.778.$

6.6 (a) All of the probabilities are between 0 and 1 and they sum to 1 so this is a legitimate probability distribution.
(b) The histogram below shows a strongly right skewed distribution with a peak at 0 days.



(c) The event $Y < 7$ is the event that the person "did not work out all 7 days."
$P(Y < 7) = 1 - P(Y = 7) = 1 - 0.02 = 0.98.$
(d) The event "worked out at least once" is the event $Y \geq 1$. $P(Y \geq 1) = 1 - P(Y = 0) = 1 - 0.68 = 0.32.$

6.7 (a) The outcomes that make up the event A are 7, 8, and 9.
$P(A) = P(X = 7) + P(X = 8) + P(X = 9) = 0.058 + 0.051 + 0.046 = 0.155.$
(b) The outcomes that make up the event B are 1, 3, 5, 7 and 9.
$P(B) = P(X = 1) + P(X = 3) + P(X = 5) + P(X = 7) + P(X = 9)$
$= 0.301 + 0.125 + 0.079 + 0.058 + 0.046 = 0.609.$

(c) The outcomes that make up the event "*A* or *B*" are 1, 3, 5, 7, 8 and 9.

$P(A \text{ or } B) = P(X=1) + P(X=3) + P(X=5) + P(X=7) + P(X=8) + P(X=9)$
$= 0.301 + 0.125 + 0.079 + 0.058 + 0.051 + 0.046 = 0.660.$
This is not the same as $P(A) + P(B)$ because *A* and *B* are not mutually exclusive. The outcomes 7 and 9 are included in both events.

6.8 (a) The outcomes that make up the event *A* are 1, 2, 3, 4, 5, 6, 7. From exercise 6.6(d), $P(Y \geq 1) = 0.32$.
(b) The outcomes that make up the event *B* are 0, 1, 2, 3, 4.
$P(B) = P(Y=0) + P(Y=1) + P(Y=2) + P(Y=3) + P(Y=4) = 0.68 + 0.05 + 0.07 + 0.08 + 0.05 = 0.93.$
(c) The event "*A* and *B*" has the outcomes 1, 2, 3, 4. So $P(A \text{ and } B) = 0.05 + 0.07 + 0.08 + 0.05 = 0.25$.
This is not the same as $P(A) \cdot P(B) = 0.32(0.93) = 0.2976$ because *A* and *B* are not independent.

6.9 (a) The net gain is either –\$1, with a probability of 0.75, or \$2, with a probability of 0.25.

X	–\$1	\$2
Probability	0.75	0.25

(b) $E(X)$ = (–\$1)(0.75) + (\$2)(0.25) = –\$0.25. If the player makes many \$1 bets, he will lose about \$0.25 per \$1 bet, on average.

6.10 The company earns \$300 with a probability of 0.9998 and earns –\$199,700 with probability 0.0002.

Y	\$300	–\$199,700
Probability	0.9998	0.0002

(b) The expected value of *Y* is $\mu_Y = (\$300)(0.9998) + (-\$199{,}700)(0.0002) = \$260$. If the company insures many, many homes, they will earn about \$260 per policy, on average.

6.11 $\mu_X = 0(0.1) + 1(0.2) + 2(0.3) + 3(0.3) + 4(0.1) = 2.1$. If many, many undergraduates performed this task, they will make about 2.1 nonword errors, on average.

6.12 $\mu_X = 0(0.03) + 1(0.16) + 2(0.30) + 3(0.23) + 4(0.17) + 5(0.11) = 2.68$. If many, many kids participated in this experiment, they would play with about 2.68 toys, on average.

6.13 (a) The mean of the random variable *Y* is located at 5 because this distribution is symmetric and 5 is located at the center. In other words, 5 is the balance point of the distribution.
(b) According to Benford's law, the expected value of the first digit *X* is

$\mu_X = 1(0.301) + 2(0.176) + 3(0.125) + 4(0.097) + 5(0.079) + 6(0.067) + 7(0.058) + 8(0.051) + 9(0.046)$
$= 3.441$. To detect a fake expense report, compute the sample mean of the first digits and see if it is closer to 5 (suggesting a fake report) or near 3.441 (consistent with a truthful report).
(c) Under the equally likely assumption, $P(Y > 6) = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{3}{9} = 0.333$. Under Benford's law

$P(X > 6) = 0.058 + 0.051 + 0.046 = 0.155$. To detect a fake expense report, compute the proportion of first digits that begin with 7, 8, or 9. A value closer to 0.333 suggests a fake report and a value closer to 0.155 is consistent with a truthful report.

6.14 (a) The company has collected 5 payments of \$250 each (for a total of \$1250) and has to pay out \$100,000. This means the company earns \$1250 – \$100,000 = –\$98,750.

(b) The missing probability is 1 – 0.00183 – 0.00186 – 0.00189 – 0.00191 – 0.00193 = 0.99058.
(c) $\mu_X = (-\$99,750)(0.00183) + (-\$99,500)(0.00186) + (-\$99,250)(0.00189)$ $+(-\$99,000)(0.00191) + (-\$98,750)(0.00193) + (\$1,250)(0.99058) = \$303.35$. If the company insures many, many people, they will make about \$303.35 per life insurance policy, on average.

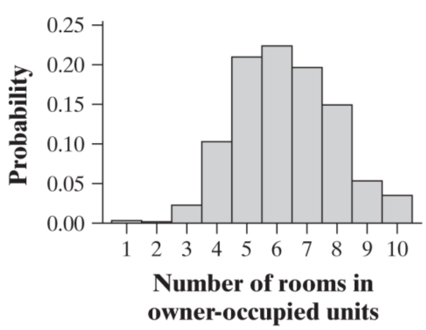
6.15 $\sigma_X^2 = (0-2.1)^2(0.1) + (1-2.1)^2(0.2) + (2-2.1)^2(0.3) + (3-2.1)^2(0.3) + (4-2.1)^2(0.1) = 1.29$. So $\sigma_X = \sqrt{1.29} = 1.1358$. The number of nonword errors in a randomly selected essay will typically differ from the mean (2.1) by about 1.14 words.

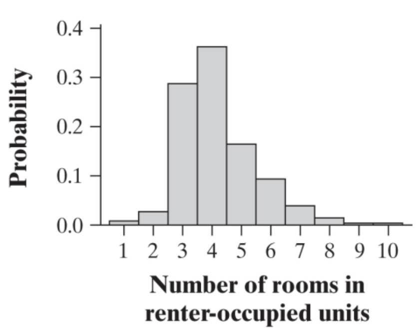
6.16 $\sigma_X^2 = (0-2.68)^2(0.03) + (1-2.68)^2(0.16) + (2-2.68)^2(0.3) + (3-2.68)^2(0.23) + (4-2.68)^2(0.17)$ $+(5-2.68)^2(0.11) = 1.7176$. So $\sigma_X = \sqrt{1.7176} = 1.3106$. The number of toys played with by a child will typically differ from the mean (2.68) by about 1.31 toys.

6.17 (a) $\sigma_Y^2 = (1-5)^2(0.111) + (2-5)^2(0.111) + \ldots + (9-5)^2(0.111) = 6.667$. So $\sigma_Y = \sqrt{6.667} = 2.58$.
(b) $\sigma_X^2 = (1-3.441)^2(0.301) + (2-3.441)^2(0.176) + \ldots + (9-3.441)^2(0.046) = 6.0605$. So $\sigma_X = \sqrt{6.0605} = 2.4618$. This would not be the best way to tell the difference between a fake and a real expense report because the standard deviations are not too different from one another.

6.18 (a) Even though the company will lose a large amount of money on a small number of policyholders who die, it will gain a small amount from many thousands of 21-year-old men. Because they insure so many men, the insurance company can expect to make \$303.35 per insurance policy, on average. The insurance company is relying on the Law of Large Numbers.
(b) $\sigma_Y^2 = (-\$99,750 - 303.35)^2(0.00183) + \ldots + (\$1,250 - 303.35)^2(0.99058) = 94,236,826.64$. So $\sigma_Y = \sqrt{94,236,826.64} = \$9,707.57$.

6.19 (a) The probability histograms are shown below. The distribution of the number of rooms is roughly symmetric for owners (graph on the left) and skewed to the right for renters (graph on the right). Renter-occupied units tend to have fewer rooms (center about 4) than owner-occupied units (center about 6). There is more variability in the number of rooms for owner-occupied units.

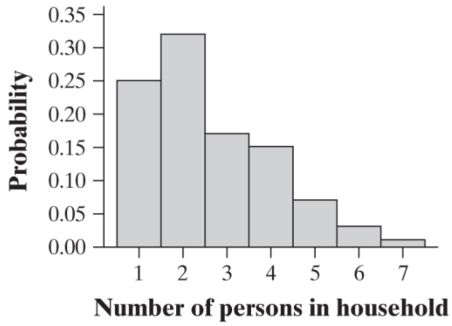


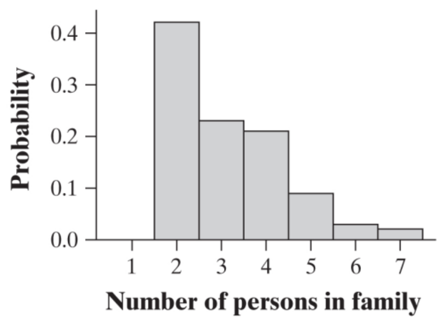


(b) The mean for owner-occupied units is μ_X = (1)(0.003) + (2)(0.002) + (3)(0.023) + (4)(0.104) + (5)(0.210) + (6)(0.224) + (7)(0.197) + (8)(0.149) + (9)(0.053) + (10)(0.035) = 6.284 rooms. The mean for renter-occupied units is μ_Y = (1)(0.008) + (2)(0.027) + (3)(0.287) + (4)(0.363) + (5)(0.164) + (6)(0.093) + (7)(0.039) + (8)(0.013) + (9)(0.003) + (10)(0.003) = 4.187 rooms. This isn't surprising because single people and younger people are more likely to rent and need less space than people with families.

(c) $\sigma_X^2 = (1 - 6.284)^2(0.003) + (2 - 6.284)^2(0.002) + (3 - 6.284)^2(0.023) + (4 - 6.284)^2(0.104) + (5 - 6.284)^2(0.210) + (6 - 6.284)^2(0.224) + (7 - 6.284)^2(0.197) + (8 - 6.284)^2(0.149) + (9 - 6.284)^2(0.053) + (10 - 6.284)^2(0.035) = 2.68934$ and $\sigma_X = \sqrt{2.68934} = 1.6399$. The number of rooms in a randomly selected owner-occupied unit will typically differ from the mean (6.284) by about 1.6399 rooms. $\sigma_Y^2 = (1 - 4.187)^2(0.008) + (2 - 4.187)^2(0.027) + (3 - 4.187)^2(0.287) + (4 - 4.187)^2(0.363) + (5 - 4.187)^2(0.164) + (6 - 4.187)^2(0.093) + (7 - 4.187)^2(0.039) + (8 - 4.187)^2(0.013) + (9 - 4.187)^2(0.003) + (10 - 4.187)^2(0.003) = 1.71003$ and $\sigma_Y = \sqrt{1.71003} = 1.3077$. The number of rooms in a randomly selected renter-occupied unit will typically differ from the mean (4.187) by about 1.3077 rooms.

6.20 (a) Both distributions are skewed to the right. The center for the "family" distribution is greater than the center for the "household" distribution, but the spread of the "family" distribution is less than the spread of the "household" distribution. Also, the event $X = 1$ has a much higher probability in the "household" distribution. This reflects the fact that a family must consist of two or more person.





(b) The means are:
$\mu_X = 1(0.25) + 2(0.32) + 3(0.17) + 4(0.15) + 5(0.07) + 6(0.03) + 7(0.01) = 2.6$ people for a household and $\mu_Y = 1(0) + 2(0.42) + 3(0.23) + 4(0.21) + 5(0.09) + 6(0.03) + 7(0.02) = 3.14$ people for a family. The family distribution has a slightly larger mean than the household distribution, matching the observation in part (a) that family sizes tend to be larger than household sizes.

(c) $\sigma_X^2 = (1 - 2.6)^2(0.25) + (2 - 2.6)^2(0.32) + (3 - 2.6)^2(0.17) + (4 - 2.6)^2(0.15) + (5 - 2.6)^2(0.07) + (6 - 2.6)^2(0.03) + (7 - 2.6)^2(0.01) = 2.02$ and $\sigma_X = \sqrt{2.02} = 1.421$. The number of people in a randomly selected household will typically differ from the mean (2.6) by about 1.421 people.

$\sigma_Y^2 = (1 - 3.14)^2(0) + (2 - 3.14)^2(0.42) + (3 - 3.14)^2(0.23) + (4 - 3.14)^2(0.21) + (5 - 3.14)^2(0.09) + (6 - 3.14)^2(0.03) + (7 - 3.14)^2(0.02) = 1.5604$ and $\sigma_Y = \sqrt{1.5604} = 1.249$. The number of people in a randomly selected family will typically differ from the mean (3.14) by about 1.249 people.

6.21 (a) $P(X > 0.49) = 0.51$.

(b) $P(X \geq 0.49) = 0.51$. *Note: (a) and (b) are the same because there is no area under the curve at any one particular point.*

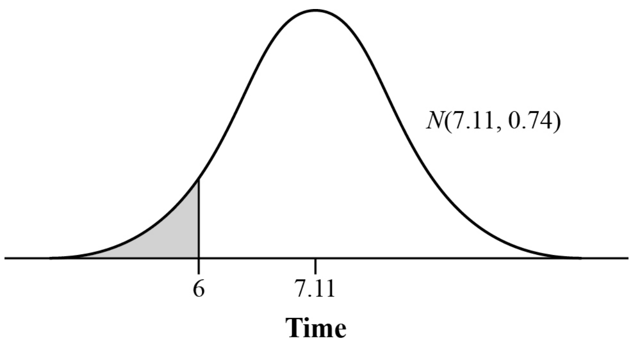
(c) $P(0.19 \leq X < 0.37 \text{ or } 0.84 < X \leq 1.27) = P(0.19 \leq X < 0.37) + P(0.84 < X \leq 1.27)$. Note that $P(0.19 \leq X < 0.37) = 0.37 - 0.19 = 0.18$. Note also that X cannot be bigger than 1, so $P(0.84 < X \leq 1.27) = P(0.84 < X \leq 1) = 1 - 0.84 = 0.16$. Therefore $P(0.19 \leq X < 0.37 \text{ or } 0.84 < X \leq 1.27) = 0.18 + 0.16 = 0.34$.

6.22 (a) $P(Y \leq 0.4) = 0.4$.

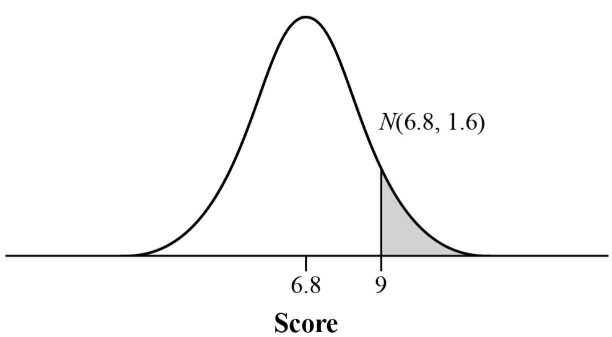
(b) $P(Y < 0.4) = 0.4$. *Note: (a) and (b) are the same because there is no area under the curve at any one particular point.*

(c) $P(0.1 < Y \leq 0.15 \text{ or } 0.77 \leq Y < 0.88) = P(0.1 < Y \leq 0.15) + P(0.77 \leq Y < 0.88)$
$= (0.15 - 0.1) + (0.88 - 0.77) = 0.05 + 0.11 = 0.16$.

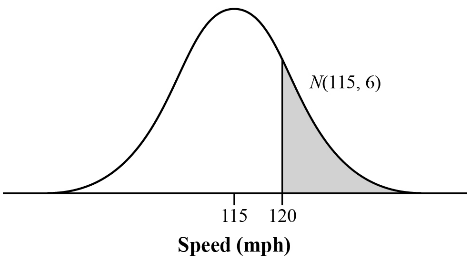
6.23 **Step 1: State the distribution and values of interest.** The time *Y* of a randomly chosen student has the *N*(7.11, 0.74) distribution. We want to find $P(Y<6)$. **Step 2: Perform calculations. Show your work.** The standardized score for the boundary value is $z=\frac{6-7.11}{0.74}=-1.50$. The desired probability is $P(Z<-1.50)=0.0668$. *Using technology:* The command normalcdf(lower: –1000, upper: 6, μ: 7.11, σ: 0.74) gives an area of 0.0668. **Step 3: Answer the question.** There is about a 7% chance that this student will run the mile in under 6 minutes.



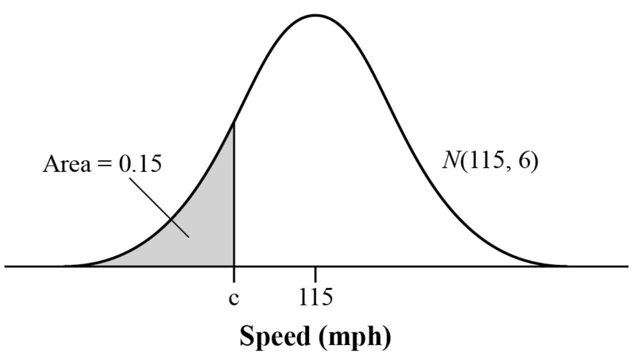
6.24 **Step 1: State the distribution and values of interest.** The score *X* of a randomly chosen student has the *N*(6.8, 1.6) distribution. We want to find $P(X \geq 9)$. **Step 2: Perform calculations. Show your work.** The standardized score for the boundary value is $z=\frac{9-6.8}{1.6}=1.38$. From Table A, the desired probability is $P(Z \geq 1.38) = 1 - 0.9162 = 0.0838$. *Using technology:* The command normalcdf(lower: 9, upper: 1000, μ: 6.8, σ: 1.6) gives an area of 0.0846. **Step 3: Answer the question.** There is a 0.0846 probability of selecting a student with a score of at least 9.



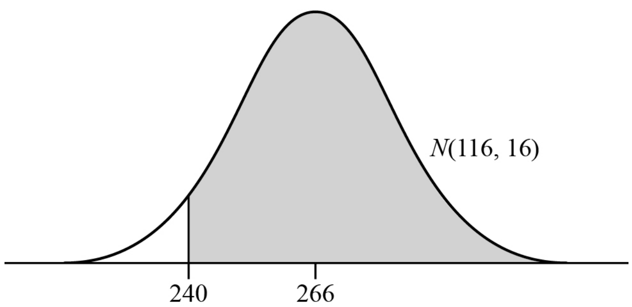
6.25 (a) **Step 1: State the distribution and values of interest.** The speed *Y* of a randomly chosen serve has the *N*(115, 6) distribution. We want to find $P(Y > 120)$. **Step 2: Perform calculations. Show your work.** The standardized score for the boundary value is $z = \frac{120-115}{6} = 0.83$. From Table A, the desired probability is $P(Z > 0.83) = 1 - 0.7967 = 0.2033$. *Using technology:* The command normalcdf(lower: 120, upper: 1000, μ: 115, σ: 6) gives an area of 0.2023. **Step 3: Answer the question.** There is a 0.2023 probability of selecting a serve that is greater than 120 mph.



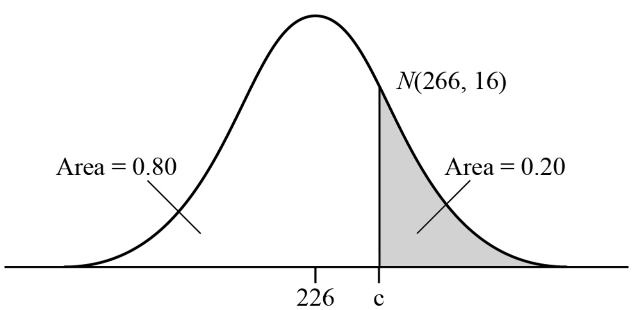
(b) The line above 120 has no area, so $P(Y \geq 120) = P(Y > 120) = 0.2023$.
(c) **Step 1: State the distribution and values of interest.** The speed *Y* of a randomly chosen serve has the *N*(115, 6) distribution. We want to find *c* such that $P(Y \leq c) = 0.15$ (see graph below). **Step 2: Perform calculations. Show your work.** Look in the body of Table A for a value closest to 0.15. A *z*-score of –1.04 gives the closest value (0.1492). Solving $-1.04 = \frac{c-115}{6}$ gives $c = 108.76$. *Using technology:* The command invNorm(area: 0.15, μ: 115, σ: 6) gives a value of 108.78. **Step 3: Answer the question.** 15% of Nadal's serves will be less than or equal to 108.78 mph.



6.26 (a) **Step 1: State the distribution and values of interest.** The length X of a randomly chosen woman's pregnancy has the $N(266, 16)$ distribution. We want to find $P(X \geq 240)$. **Step 2: Perform calculations. Show your work.** The standardized score for the boundary value is $z = \frac{240-266}{16} = -1.63$. From Table A, the desired probability is $P(Z \geq -1.63) = 1 - 0.0516 = 0.9484$. *Using technology:* The command normalcdf(lower: 240, upper: 1000, μ: 266, σ: 16) gives an area of 0.9479. **Step 3: Answer the question.** There is a 0.9479 probability of selecting a woman whose pregnancy lasts at least 240 days.



(b) The line above 240 has no area, so $P(X \geq 240) = P(X > 240) = 0.9479$.
(c) **Step 1: State the distribution and values of interest.** The length X of a randomly chosen woman's pregnancy has the $N(266, 16)$ distribution. We want to c such that $P(X \geq c) = 0.20$ (see graph below). **Step 2: Perform calculations. Show your work.** Look in the body of Table A for a value closest to 0.80. A z-score of 0.84 gives the closest value (0.7995). Solving $0.84 = \frac{c-266}{16}$ gives $c = 279.44$. *Using technology:* The command invNorm(area: 0.80, μ: 266, σ: 16) gives a value of 279.47. **Step 3: Answer the question.** 20% of pregnancies last at least 279.47 days.



6.27 b

6.28 b

6.29 c

6.30 c

6.31 Yes, in general, students did have higher scores after participating in the chess program. If we look at the differences in the scores (post – pre), the mean difference was 5.38 and the median difference was 3. This means that at least half of the students (though less than three quarters because Q_1 was negative) improved their reading scores.

6.32 No, we cannot conclude that chess causes an increase in reading scores. We do not have a control group that did not participate in the chess program for comparison. It may be that children of this age improve their reading scores for other reasons (e.g., regular school) and that the chess program had nothing to do with their improvement.

6.33 The equation of the linear regression model is:
$\text{predicted posttest} = 17.897 + 0.78301(\text{pretest})$.

6.34 This linear model is appropriate because the residual plot does not show any leftover patterns. 55.8% of the variation in posttest score is accounted for by this linear model relating posttest score to pretest score. Finally, when using this model to predict posttest scores from pretest scores, the actual values will typically be about 12.55 from the predicted values.